

FUNDAMENTAL CHARACTERISTICS OF THE DICTIONARY AND GRAMMATICAL SCHEMES USED IN MACHINE TRANSLATION FROM ENGLISH INTO RUSSIAN

By I. K. BEL'SKAYA

Machine Translation, pp. 47-80

The object of the present communication is to discuss work carried out by the author and his colleagues in the machine translation (MT) group at ITM and VT (Inst. of Precision Mechanics and Computer Technique) of the USSR, G. A. Tarasova, A. I. Martynova, L. M. Bykova and T. M. Nikolayeva, on carrying out MT from English into Russian.

At the present time we are approaching the completion of the second stage of the work, which is concerned with the extension of the principles of MT put forward in the first stage* to new fields (particularly other languages) and also study of the details and a certain generalisation of these in the old field, *i.e.*, English-Russian MT.

The advances achieved in the second stage consist primarily in extending the Anglo-Russian MT dictionary to more than double its volume. The English section alone now contains 2,300 words and makes possible adequate translation of scientific literature on different problems in applied mathematics.

In compiling and checking it the following works by English authors were used:

1. Milne, E. Numerical solution of differential equations.
2. Karman and Biot. Mathematical methods in engineering.
3. Salvadori and Baron. Numerical methods in engineering.
4. Hardy. Divergent series, Oxf., 1949.
5. Bellman. Stability theory of differential equations, New York, Toronto, London, 1953.
6. McLachlan. Theory and application of Mathiew Functions, Oxf., 1947.

As regards the grammatical schemes for translation, these schemes at present provide for analysis of all English grammatical constructions and also word-variations ("synthesis" in our terminology) of any Russian word form. Only a few types of inversion which are not characteristic of the style of scientific technical literature, are provided by English analysing schemes.

The completion of the second stage of the investigations on the construction of an Anglo-Russian system of MT consisted in the "manual" form of MT, *i.e.*, extended test checking of the dictionary and the translating schemes by means of a group of ten people who did not know English and who therefore carried out the whole cycle of operations prescribed by the MT schemes entirely mechanically. About 3,000 phrases of "unknown text" consisting of more than 100 extracts from the texts of different authors† were translated by way of test.

The test checking gave positive results, showing the suitability of the translation schemes and the adequacy of the scope of the 2,300-word dictionary for satisfactory translation of English literature on applied mathematics into Russian. The number of "unknown words" with the present size of the dictionary did not exceed on the average 1-2 words per page. We therefore consider it possible to put off expanding the mathematical dictionary further until such time as mass translation of literature by means of a machine commences.

We do not propose to expand our "manual" dictionary for the further reason that we do not want to deprive it of its present specialisation for we consider that the narrower the specialisation the more it corresponds to its basic function: cutting down multiple meanings of words. Further extension of the MT dictionary is therefore understood by us as the creation of a large number of mutually unconnected specialised dictionaries ("micro-dictionaries"). It is precisely in this

* We regard the end of the first stage of the work on MT at ITM and VT as being the well-known experiment in translation from English using the BESM machine (1955).

† Several extracts from the texts so translated are given as an appendix to the present paper.

direction that we propose to carry out our work in the third stage of the research on Anglo-Russian MT.

As regards the translation scheme and the general construction of a system of Anglo-Russian MT we have come to the conclusion that it is possible and desirable to keep them in the present form and state.

The construction of the MT system is such that the whole process of translation can be divided into three main independent parts:

1. Dictionary analysis of English phrases—
by means of the dictionary and dictionary schemes.
2. Grammatical analysis of English phrases—
by means of grammatical analysis schemes.
3. Grammatical synthesis of Russian phrases—
by means of Russian synthesis schemes used indifferently for grammatical agreement of phrases translated from different languages.

I. Dictionary Analysis of the English Phrase

Dictionary analysis begins with a search for the words in the phrase to be translated in the Anglo-Russian MT dictionary by the method of applying the words of the phrase to the dictionary and provision is made where necessary for transforming the word form in the context to the dictionary form ("scheme for finding words in the dictionary").

For words which cannot be found ("unknown words") analysis is carried out to ascertain which of nine lexicological-grammatical classes—parts of speech—includes them ("scheme for determining part of speech of 'unknown words'"). We attach great importance to this scheme since it enables "blanks" to be eliminated in a sentence and a reliable grammatical analysis to be made even in the case where the dictionary does not give 100 per cent. word coverage in all fields which as we know is impossible.

Besides "unknown words", so-called "formulae" are also subjected to preliminary grammatical analysis, *i.e.*, conventional notation of every kind which are not words and therefore not included in the dictionary ("Scheme for analysis of 'formulae'"). The necessity of analysing these is dictated by the diversity of their functions in a sentence. The object of the analysis is final syntactic classification ("attestation": this word is often used by Russians in senses like "examining" (committee) suggesting borrowing of the Latin root without proper understanding of its meaning—*Tr.*) of all "formulae" in a phrase by attributing features to them which place them in the same class as the other components of the phrase-words. After the operation of this scheme each formula has one of the following four classifications ("attestations"!).

- (1) Noun with the feature "sing. no." or "plur. no." sometimes with the feature "genitive case".
- (2) Adjective with feature "ordinal".
- (3) Verb with features "predicate", "invariable".
- (4) AP (absolute passage).

The feature "noun, gen. case", is given to a formula appearing in the function of determination to a noun and in preposition to this noun. A noun can have an analogous position in an English phrase in the function of determination to another noun and in this case a Russian translation of the English combination will correspond to a transference of the determination to the post-position after giving it the feature "genitive case". Giving analogous features to a formula unifies this operation:

Ex. *X* variable—переменная (variable (of) *X*).
Coefficient functions—функции коэффициентов (functions of the coefficient).

In the remaining cases a formula with the feature "noun" has no case indication in this scheme, but for this purpose it is referred to the scheme for analysing nouns.

The feature "adjective, ordinal" are given to formulae having the termination "-th"

e.g., i-th term → i-тый член (ith term).

N-th partial sum → N-ая частная сумма (Nth partial sum).

The feature "verb, predicate" is an indication that in its structure and position in the phrase the formula is equivalent to a whole sentence the main feature of which is regarded as its predicative nature (in the formula it may be expressed by signs =, ≠, <, >, →).

Test 714 (RB p.61-62)

The	system	has	the	form	(14) $\frac{dz}{dt} = \left(A_0 + \frac{A_1}{t} + \dots \right)$
Adj., pron., def	noun fem. gen.	verb [?] act trans. pres. sing. 3rd pers.	Adj. pron. def.	noun m. gen.	formula [AP]
	sing. nom. subj.	takes acc. case		sing. acc.	
	СИСТЕМА	ИМЕЕТ		ВИД	(14) $\frac{dz}{dt} = \left(A_0 + \frac{A_1}{t} + \dots \right)$

) (where	$\ A_2(t)\ \leq c_1 t^{-2}$	as	$t \rightarrow \infty$	
?[ZN] Punct. mk.	?[CH] adv.	formula verb trans.	adv. complex	formula verb trans.	full stop [KP]
		unchanged		unchanged	
	ГДЕ	$\ A_2(t)\ \leq c_1 t^{-2}$	По мере того, как	$t \rightarrow \infty$	

Fig. 1.

Finally the sign AP denotes that the "formula" is without syntactical significance because it may be ignored in the grammatical analysis of the phrase. "Formulae" in postposition to a defined English noun are graded in this way.

Ex. . . . there exists a matrix T —существует матрица T (RB, p. 20).

Cf. also Fig. 1.

A word found in the dictionary has a grammatical and a certain semantic characteristic in the dictionary in the form of a combination of features ("invariant characteristics of the word") and we distinguish between English and Russian invariant characteristics, for in the latter we separate specially the "analytical" (i.e., used in Analysis schemes) and "synthetic" (i.e., used only in the schemes of the synthesis) parts.

Thus before the grammatical schemes are operated we have initial information of a grammatical character for each word and each symbol ("formula") in the phrase. It is this information which is used in the Analysis schemes. Lexicological information on the word, however (its translation), is not used till the third stage of translation—the operation of the synthesis schemes.

Therefore in programming, the dictionary equivalent is not selected in the first stage but its number only is established and the accompanying information selected (analytical invariant characteristics).

It is nevertheless important that dictionary analysis of the phrase including also "selection" of the dictionary equivalent should precede grammatical analysis for otherwise we shall be unable to establish the correct original invariant characteristic of the words in the phrase, in many cases. Thus for the equivalent "follow" the adverbial characteristic (as + follows — следующим образом) can only be established after it has undergone dictionary analysis (e.g., also the suggested

characteristic of the equivalent "compare" in the position noun + compared + with — "по сравнению с" or the adjectival characteristic question in the position noun + in + question "-рассматриваемый").

Dictionary information on a word in the dictionary may be final or preliminary.

The establishment of the Russian equivalent of the word is regarded as final dictionary information; preliminary information consists in markings "homonym" or "many meanings".

A word with the marking "homonym" is treated by means of the dictionary scheme "Distinction of homonyms" which gives a solution for the following four types of homonym:

- (1) Adjective-noun (corresponding to marking "OM 1"), there are 96 words in our dictionary and some of these are included in the group "OM 1 complex" which denotes compounding of the simple (two-way) homonym:

e.g., check	—1. adjective	контрольный	(checking)
	2. noun	контроль	(a check)
	3. verb	"многозн"	(multi-meaning)
square	—1. adjective	квадратный	(square)
	2. noun	"многозн"	(multi-meaning)
	3. verb	возвести в квадрат	(bring into square)

- (2) Noun-verb (60 words in our dictionary are marked OM 2).
 (3) Verb-adjective (18 words in our dictionary are marked OM 3).
 (4) Adverb-preposition (20 words in our dictionary are marked OM 4 since this number does not include words requiring individualised analysis).

The delimitation of the first three types of homonym is carried out by a combination of morphological and syntactic analysis of a word with the marking "homonym". Thus the presence of any ending (except -er or -est) with a OM 1 or OM 3 immediately excludes the alternative "adjective" just as the presence of the ending -ing or -ed in OM 2 solves the question in the favour of the verb. For all remaining cases syntactic analysis is used.

As a result of the scheme "Differentiation of homonyms" part of the words are given final dictionary information and part only of the marking "multiple-meaning".

All words with the marking "multiple meaning" (their number in our dictionary is 1/5th of the total number of words in the dictionary, i.e., 458 words) are treated under the scheme "Analysis of multiple meaning" as a result of which all ("known") words without exception in a phrase obtain final dictionary information accompanied by the original invariant characteristics.

The scheme "analysis of multiple meanings of words" provides contextual analysis of words with multiple meanings and relies predominantly on "micro-contexts" but taking the "macrocontext" into account (in the interpretation which we suggested in the general paper "Some general principles of MT").

A few examples of analysis on the basis of the "microcontext":

. . . . 491 design (I, a): check the nearest noun on the right (*—2^a).

("OM 2" → verb, "mult. meaning")/if "no" then to the left (*—2^b) to the symbol ("H-Pr") ("purpose of apparatus") or group METHOD.

a(Π, Ш) Check nearest preposition on left for sign ("PPZ") or PD ("PN") (preposition of direction) or PO ("PTs") (preposition of object).

I (0, 0) WORK OUT (verb, takes acc. case).

II (0, 0) DESTINE (verb, takes acc. case).

III (0, 0) DENOTE (verb, takes acc. case).

. . . . 48 advance (a, b): Check given word on feature "verb".

1 "OM 2" → 1. Noun "mult. meaning" *m(b, I)*: Check given word on (reject) ending -ed and following word on "noun".

b(Π, Ш): Check noun found on group TEXT-BOOK.

- (IV, V): Check preceding word for group "IN".
- I(0, 0): RAZRABOTAT' (verb taking acc. case).
- II(0, 0): VTORAYA STUPEN' (adj. and noun) f. gen. sing. determinative past par.
- III(0, 0): Completed (adject.).
- IV(0, 0): ZARANEYE (adv. of time).
- (IN —particle, not translated).
- V(0, 0): USPEKH (noun, m. gen.).

Fig. 2.

By way of example of the effect of the "macrotext" on the system of meanings of a word we give an analysis of the word "picture" in a mathematical dictionary:

1421 *picture* (1.П): check previous word on "adjective"

a-noun

I(0, 0): KARTINA (noun, f. gen.)

II(0, 0): CHERTEZH (noun, m. gen.)

In all cases it is important to note that the meanings given by the MT dictionary are "relative meanings" in the sense outlined in our paper "Some general principles of MT".

II. Grammatical Analysis of an English Phrase

Grammatical analysis is carried out by means of a cycle of special schemes (so called Analysis Schemes) six in number, the order of operating which is indicated in the headings of the present paper.

Analysis of an English phrase begins with the scheme "verb" which is the key scheme of the first half of the Analysis (up to the schemes for syntactical analysis inclusive) while syntactical analysis of the phrase (Scheme 3) gives the basis for the working of the second half of the analysis by establishing the limits of the sections within which it is carried out.

These sections are:

- (a) Simple sentence (after removing from it all separated and introductory groups of words).
- (b) Separated groups of words within the given sentence.
- (c) Introductory groups of words within the given sentence.

The sequence in which they are treated corresponds to the order in which they are enumerated.

Scheme of Analysis of verbs which operates first has at its disposal only initial dictionary information. The object of the scheme is to determine all verbal forms in the phrase. Here the verbs in the personal form are given the syntactical feature "predicate" and also features of mood, voice and tense; the features person and number at this stage are provisional in character, reflecting the English characteristic of the word and therefore later submitted to correction.

It is very important to note that from the analysis of verbs in English we obtain information of a double character: English and Russian.

We rarely fix the first (except such features as person, number and sometimes "syntactical function") while we record the second at the earliest opportunity.

Thus by analysing non-personal forms of English verbs ending in -ing we give them the feature "verbal noun", "participle", "participial adverb" or "indeterminate form" corresponding with the role of their equivalents in the Russian sentence (Cf. RB, p. 12).

By converting it into . . . verbal noun S POMOSH'YU PREEBRAZOVANIYA YEGO V . . . ("by the help of converting of it into")

Directly setting U adv. part. NEPOSREDSTVENNO POLAGAYA U
 Before turning to this "-indet. f." PREZHDE CHEM OBRATIT'SYAK

Another example. By analysing a construction with an impersonal form of verb (infinitive, participle and gerundive) we convert these constructions in the Russian interlinear translation into the final clauses and impart to the (Russian) verbs the feature of personal form—"predicate" tense, mood, &c. In translation, one impersonal form may be replaced by another: an infinitive by a verbal noun, a participle by an infinitive, &c. (see Fig. 3).

Corresponding with the special features of Russian verb constructions verb-predicates may receive the marking "subject inverts into direct (or indirect) object" (see Fig. 4).

Test 69 (RB, p. 2)

It	is	easily	verified),(that	(3) x + y ≤
subst. pron. 3rd p. sing. neut.	verb. I pred. pres. 3rd p. sing.	adj. adv.	verb takes acc.	[ZN]	[SN]	formula verb. pred.
			indef. form pred. pass through *-I	punct. mark		not changed
		ЛЕГКО	ПРОВЕРИТЬ	,	ЧТО	(3) x + y ≤

Fig. 3.

Test 62 (RB, p. 2)

Let	us	now	consider	various	operations),(
particle adj. aux.	pronoun 1st pers. plur.	adv. of time	verb takes acc.	adj.	noun f. gen.	
			predic. fut. 1st pers. plur.	f. pl. acc.	pl. acc.	[ZN] [punct.]
		ТЕПЕРЬ	РАССМОТРИМ	РАЗЛИЧНЫЕ	ОПЕРАЦИИ	

which	may	be	performed	upon	vectors
conj. adj.	verb modal pred.	verb I	verb takes acc. case	prepos. takes instr.	noun m. instr. case
→ noun, conj. subj. acc. → pl.	impers. pass by (*-I)	pred. inf. pass by (*-I)	indef. pred. "invert subj. into dir. obj."	—	pl. instr.
КОТОРЫЕ	МОЖНО		ВЫПОЛНИТЬ	НАД	ВЕКТОРАМИ

Fig. 4.

It is essential to note that analysis of verbs depends on the special structural classification characterising verbs in two planes:

- In the plane of English itself—by uniting by so called "additional" signs: "declarative" verbs (call, label, declare, report, &c.), "causative" (cause, enable, make, order, command, &c.), "modal" and "semi-modal" (help, dare), "verbs requiring two objects" (give, offer, permit, &c.), "semi-auxiliary" (seven types distinguished), &c.
- On the plane of comparison with the translating norms of Russian, by performing on so called groups (53 groups in all are distinguished, divided into three classes) so constructed to exclude any overlapping. In

the groups words are included which have special features in translation by personal forms of verbs (Class I), impersonal forms (Class II) and certain more complex cases (Class III).

After working on the scheme "verb" we obtain final information on the grammatical form of all the verbs in the Russian phrase.

Scheme of analysis of punctuation signs follows after the scheme "verb"; and uses the results of the latter. The object of the scheme is to ensure that all the punctuation marks in the English phrase (apart from the "final" ones, full stops, exclamation and question marks) which reflect firstly, the function of the sign in question in grammatical analysis of the English phrase, *i.e.*, its "English function" and secondly, the fate of the sign in translating the phrase into Russian (*i.e.*, its "Russian function"). If these functions coincide a common feature corresponds to both (Cf. the characteristics of brackets, semicolons), but if the "Russian" function differs from the English, the sign receives some additional feature. For example, commas separating totally differentiated sentence groups or adverbs in the English phrase, are given two features: English ZV (introductory sign) and the Russian ZPr (the sign pass through Russian). Russian and English signs also differ in the case where the English sign should be transferred into the Russian text, although its presence in the English phrase impedes unification of analysis. In this case, however, instead of two different features the sign is given one feature —ZPa (comma-pass through English). In this way the omission of the sign in the English analysis and its retention in the Russian translation is ensured.

The features which punctuation signs may obtain as a result of analysis of the English phrase can be summarised by means of the five main and nine supplementary features:

1. Uniform (ZO).
2. Non-uniform (ZN).
3. Introductory (ZV).
4. Separative (ZO).
5. Pass through (ZP).

Features supplementary to the first are:

- (a) Generalised-uniform (ZOo); to the second
- (b) "Small point" (MT); and to the others
- (c) General pass through (ZPo)
- (d) Pass through in English (ZPa)
- (e) Pass through in Russian (ZPr)

In addition to this paired or "separating" punctuation marks obtain additional features "beginning" and "end".

All these features are extremely important landmarks in the subsequent syntactic joining up of the phrase from (simple) sentences and determining accurately introductory and separative groups in the sentence ("*Scheme of syntactic analysis of the phrase*").

The second half of the Analysis begins with the scheme "noun and numeral" and here accuracy of analysis is increased by using information on the syntactical structure of the phrase.

Functionally numerals* are not differentiated from nouns and can be analysed by the same rules although the difference in morphological character between English nouns and numerals prevents them from being combined in one class.

The scheme of analysis of adjectives is the last fundamental English scheme with results of material importance for Russian schemes of synthesis. The scheme depends on the results of all the preceding schemes and provides all the adjectives in a Russian phrase with a variant characteristic, *i.e.*, features of gender, number, case, degree of comparison, form (long or short).

In addition the scheme covers cases of substantivation of adjectives and also the possibility of adjectivisation of numerals.

The scheme "*Change of word order in translated phrase*" completes the work on the schemes of Analysis. By means of this scheme, cases of "un-Russian" word order due to certain special features of English norms in comparison with Russian, are eliminated.

* Especially as we regard only quantitative numerals as such, "ordinals" being regarded as adjectives.

It should be noted that in translating from English to Russian, changes in word order are in the majority of cases local in nature.

The most important differences in the position of syntactic units as between Russian and English sentences are due to (a) the prepositive position of nouns in the function of definition, in respect to another noun; (b) the expression and position of negatives in Russian and English sentences.

Ex. 1. "Use the *Lagrange variation-of-parameters* method to solve the inhomogeneous n^{th} order differential equation (RB, p. 12) (=Use the method of *variation of parameters of Lagrange* to solve the differential equation of the n^{th} order).

Ex. 2. "Hence if $f(0)=0$ no nonidentically vanishing solution of the equation can have (RB, p. 74) (consequently if $f(0)=0$ no (adj. form) nonidentically vanishing solution of the equation can *not* have (double negative in Russian).

III. Grammatical Synthesis of the Russian Phrase

We term the system of rules enabling sentences translated from (another) language to be built up grammatically. Synthesis schemes.

Multi-language machine translation (*i.e.*, translation from many languages) imposes certain requirements on Synthesis schemes:

- (1) These schemes should be *exhaustive* in setting-out the rules of word variation in Russian since the grammatical rules which are not very relevant in translating from one language, may be extremely relevant in translating from another language.
- (2) These schemes should be *universal*, *i.e.*, should give a positive answer to any requirements regarding word variations arising from the Analysis schemes.
- (3) These schemes should not be dependent on Analysis schemes for the latter are of different importance for different languages.

For schemes of Synthesis, indication of "non-use" of particular grammatical categories which may be encountered in normal grammar cannot be regarded as final: it should be accompanied by an indication as to what grammatical form should replace it. The problem of the equivalence of grammatical forms of an individual language (Russian) arises and a solution is extremely important in constructing schemes of Synthesis in machine translation.

The importance of the problem of grammatical equivalence of forms of an individual language is due to the difference in the degree of productivity of equivalent grammatical categories in different languages. (Cf. the category of the verbal noun in English and Russian).

In regard to the independence of Synthesis schemes from those of Analysis we should note that it is here that one of the fundamental differences between the two (classes) of schemes lies.

Analysis schemes are comparative by nature, this being justified as follows: grammatical analysis without reference to a particular language for translation or some rigidly limited system of grammatical categories (also the actual formulation of certain conventional languages) would be too "uneconomic".

On the other hand, Synthesis schemes may be compiled in such a way as to be specialised expositions of the rules of word-variation (and certain laws of word-formation) of an individual language.

At the present time an attempt is being made to compile the following Russian Synthesis schemes for machine translation:

- (1) "Word-forming scheme" (auxiliary).
- (2) "Verb."
- (3) "Adjective."
- (4) "Noun and numeral."

Testing of a version of these schemes programmed for a machine, has given satisfactory results: the machine constructed without error the required word forms of the words of a Russian interlinear translation. The Russian sentences worked out by the Synthesis schemes which sometimes lacked stylistic finish, were unexceptionable from the grammatical point of view.

I. Word-forming Scheme

Among the alternative (contextual) grammatical features which a Russian word acquires in the Analysis schemes, word-changing and word-forming are differentiated.

By *word-forming features* we mean those which require consultation of the dictionary for the purpose of selecting a new (dictionary) word; *word-varying features* are obtained by simple change in the termination of the word in accordance with rules combined in the basic Synthesis schemes.

We have thus, for our work, modified the meaning of the terms word-change and word formation: any transformation of a word which does not involve the disappearance of the root of the word is regarded by us as word-variation: on the other hand any replacement of a word by a new (dictionary) word (including suppletive form formation) is considered as the formation of a new (dictionary) word—word-formation.

"The word-forming scheme" is divided into four independent parts—corresponding with the main variable classes of Russian words.

In the first part are included the adjectives which in the schemes of Analysis, received the features "adverbial", "substantival" or "degree of comparison—not positive." In addition it includes adjectives forming the short form irregularly (e.g., BOLSHOY-VELIK) or plural (some pronomial adjective).

The feature, "adverbial adjective" occurs as "word-forming" or as "word-changing" depending on the type of formation of the adverbial form of the different groups of adjectives in Russian. This feature is word-formative for pronomial adjectives and also for adjectives not forming the adverbial form synthetically.

Among adjectives having an analytic adverbial construction in Russian, four types are distinguished:

1. Adjectives corresponding to the construction "by means of + a noun of one group with the adjective in question".
2. Adjectives corresponding to the construction "in the form of + a noun of one group with the adjective in question".
3. Adjectives corresponding to the construction "of + a noun of one group with the adjective in question".
4. Adjectives corresponding to the construction "by way of + a noun of one group with the adjective in question".

In all cases the noun taken from the dictionary receives the feature "gen. sing". The sign "adjective" hitherto accompanying the word is deleted. The special features in carrying out the adverbial function of the adjective are fixed in the form of the dictionary feature of the group of the adjective (groups Iap-N, Ibp-N, Icp-N and Idp-N*).

For the remaining groups of adjectives, the sign "adverbial" is not word-formative.

The feature "substantival" which may be acquired by pronomial adjectives is always word-formative.

Ex. (English) The approximation becomes invalid because of *our* giving the variable a wrong value.

(Russian: The approximation becomes invalid because *we* give the variable a wrong value).

== Приближение становится бессмысленным, потому что *мы* даем переменной неверное значение

* The notation of each group consists of the type number (roman numeral), serial number of the group (alphabetical letter following roman numeral), indication of the part of speech (p in the foregoing example means adjective) and the name of the group ["detachment"] (in the example N means: adverbial construction formed analytically).

In this case the adjectives are checked for dictionary features of person, number and gender and are replaced by personal pronouns with analogous dictionary features:

E.g., my—I, his—he, thy—thou, &c.

The features of comparative and superlative degrees of comparison of an adjective in the majority of cases* are also word-formative since they are obtained by the addition of a new word to the text: where the feature “comparative degree” occurs the word “более” is taken from the dictionary and where “superlative degree” the word “наиболее” is added to the existing adjective.

Of the two methods of forming the degrees of comparison in Russian—synthetic and analytical—the latter is preferred as it is more universal. The synthetic method is used for the formation of degrees of comparison from adjectives in the short or adverbial form only and of a small group of irregular adjectives (good, bad, big, little, small). In a number of cases, particularly for the above group of irregular adjectives, the adjective occurring in the text is replaced by the word from the Russian dictionary corresponding to the alternative features of the adjective in question:

E.g., KHOROSHY—LUCHSHIY (good—best).
PLOKHOY—KHUDSHY (bad—worst).
PLOKHOY (+“adv. form”)—KHUZHE (worse).

In the second part of the “word-forming scheme” nouns having received the feature “adjectival” in the Analysis schemes are dealt with; this indicating that this adjective should be replaced by an adjective taken from the dictionary. Personal pronouns may receive this feature on translation from Chinese in particular:

根據我們國家的情況提出了這個問題。

(this question arises in connexion with the position of our country).

Apart from this the “word-forming scheme” gives features “plural number” for nouns forming the plural in a non-standard manner:

E.g., CHELOVEK—LYUDI; GOD—LET (plur. and gen.).

Cf. also such formation of the plural of pronouns in translation from Chinese, Japanese and similar languages:

在證明 $n=2$ 時我們需要以下的二引。

(in proving $n=2$ we require the following two lemmas).

In the third part of the “word-forming scheme” numerals are dealt with.

In the Analysis schemes, numerals acquire the following word-forming features: “introductory”, “ordinal” and “substantive”.† The first feature is obtained by prefixing on the left, the particle V/O/— the features “introductory” being replaced by “ordinal” and the affixing of the additional feature “plural no., prepositional case”, after which the word-varying scheme ensures correct transformation of the word.

The second feature “ordinal”‡ is obtained differently for numerals with the feature “less than five” and those without it: in the first case we select from the dictionary the word corresponding to that given as regards number and feature; in the second case the ending -OY or -Y is added to the word given, depending on the dictionary group of the numeral. Possible changes in the root of the word are also fixed by means of dictionary features. In both cases the transformation is accompanied by a change in the basic dictionary feature: from “numeral” it becomes “adjective”.

* Exceptions are adjectives in the short or adverbial form, the degree of comparison being formed synthetically from the same root of the word.

† Such a system of characteristics of the numeral enables us to reduce to $\frac{1}{2}$ the volume of the section of the dictionary containing numerals: the words naming the numbers of the natural series (1, 2, 3, &c.) are given in the dictionary once only and not repeated for the ordinals (1st, 2nd, 3rd) or “introductory” in words (firstly, secondly), or “substantivising” (a set of 4, 5, 6, &c.). Of these the dictionary only includes some cases of “irregular” word formation.

‡ Russian numerals receive this feature in Analysis schemes in cases where the equivalent numeral in the original has no morphological form enabling it to be characterised as a “quantitative adjective” before applying the Analysis scheme. Cf. Chinese scheme for numerals.

The feature "substantive" is taken by numerals where it is required to transform them into the form two, three, four, &c. For some numerals this is carried out by resorting to the dictionary, but in the majority of cases it is sufficient to add the required ending—according to the established rules.

In the fourth part of the "word-forming scheme of synthesis" verbs are varied with the feature "subjunctive mood" or "future tense", imperfective form. In the first case, the additional word "БЫ" (with the feature "particle") is selected from the dictionary and in the characteristics of the verb in question the feature "subjunctive mood" is altered to "past tense", and the verb is then passed to the word-variation scheme of synthesis.

In the second case the verb БЫТЬ (to be) is also taken from the dictionary and the feature "future time, perfective form" is given to it while the verb in the text receives the feature "indeterminate form".

In addition to this, the fourth part of the word-forming scheme of synthesis solves the problem of "non-standard formation" * of such verb categories as "imperfective form", "verbal noun", "participle" and "adverbial participle" when the feature is found by consulting the dictionary.

Thus "adverbial participle of imperfective form" from the verb SKAZAT' is formed in the word-variation scheme only after the verb has passed through the word-forming scheme, since in this case the original verb should be replaced by another word—GOVORIT' and the required form is obtained from it by word-variation.

II. Main Schemes of Synthesis

1. The scheme "verb" is the first scheme in the group of main schemes of Synthesis. It consists of two sub-schemes ("use of forms of verb" and "formation of verb forms") which are of a preliminary nature and of a specifically word-variation verb scheme.

In the preliminary sub-schemes, the category of the form of verbs in the translated phrase is found exactly by contextual analysis of the Russian phrase and after this the verbs are passed to the main word-variation scheme.

In the working of the verb word-variation scheme four stages are distinguished:

In the first stage a system of checking determines exactly the group of words which should be changed according to the scheme in question: verbs are selected not having the feature "invariable" or the combination of the two features "indeterminate form and active voice".

In the second stage the root of the verb is established, to which the required ending is subsequently added. The first step in this direction is to reduce the dictionary endings of the verbs to a single form: the particle -SYA (indicates reflexive or passive in Russian) if the dictionary form of the verb terminates in this way. Then the one of the two roots of each verb which corresponds to the alternative features of that verb is obtained.†

—*I root (present)* is obtained where the verb has one of the features "present tense", "future tense", "imperative mood" or the combination of two features "adverbial participle and imperfective form".

—*II root (infinitive)* is obtained in all other cases, i.e., where there is the feature "past tense" or "verbal noun" or the combination of the two features "adverbial participle + perfective form".

In all cases typical root formation is preceded by separation of group or single deviations from the normal formation. The deviation is determined by means of the dictionary feature of the group of the verb.

* "Non-standard formation" is the name given to all cases of "non-use" of particular grammatical categories in the language of the translation in which the necessity arises of substituting their equivalents for the unused categories.

† An exception is verbs in -Ch' and -TI (PECH', UVLECH', MESTI, NESTI, &c.) except -RASTI, IDTI and verbs ending in -YTI). These verbs are separated into a group of "single root" verbs since all the grammatical forms of these verbs which concern us are formed from the same root.

By normative formation of the I root of the verb is meant that in which the two last letters are removed from the dictionary form* for verbs in the first conjugation or the last three letters in second conjugation verbs, the remaining part of the word being retained without any alteration in the root of the word. All cases of alteration of the root are deviations from the normal. Among "irregular" verbs we differentiate those which vary one, two or three final letters,† beyond this "irregular" verbs are combined in groups corresponding to the letter (or combination of letters) which appear in the root of the verb as a result of alteration.

The normal (normative) formation of the II root of a verb consists in eliminating the last two letters from the dictionary form of the verb after its unification in respect to the particle -SYA. Deviations from the above rule are taken account of for the following three groups of verbs: (1) the verb RASTI and verbs derived from it; (2) the verb IDTI and verbs ending in -ITI and (3) verbs reducing the root to (by?) one letter.

In the III Stage of operation of the verb scheme, one of two cycles of further checking is followed: *Cycle A* (for I root) with further determination of the following categories:

1. "Participle" → 2. "Adverbial participle" → 3. "Imperative mood" → 4. (Final "No.") "present tense" or "future tense (perfective form)".

Cycle B (for II root) with further determination of the following categories:

1. "Participle" → 2. "Adverbial participle" → 3. "Verbal noun" → 4. (Final "No.") "past tense".

In Cycle A are included two participial forms of the present tense—active and passive and one adverbial participial form of the verb—the adverbial participle of the imperfective form. In Cycle B are included the participial forms of the past tense and the adverbial participle of the perfective form. Verbal nouns and participles, categories which are double in essence are worked out in the scheme "verb" only in the form of the root, being subsequently transferred to the scheme "adjective" and "noun" respectively for variation as regards case and number.

Within the limits of the category of Cycle A or Cycle B, which is of importance for the verb in question, the form is further determined: for "participle" the voice is determined, for "present tense" and "future tense" the person and number and for "past tense" the gender and number are fixed. Thus at this stage of the working of the scheme it is made clear what ending should be added to the verb root. For some groups of verbs‡ analysis in respect of category also determines a number of particular variations in the root of the word. Thus for groups I Kg-NV§ account is taken of alteration in the root in the first person sing., present tense (NOCIT'—NOSHU); for groups I ag-PV, the special feature in forming the past tense is taken into account (ZAMERZNUT'—ZAMERZ); for group I bg-PP the special feature in the formation of the past participle (use of the suffix -T), &c., is allowed for.

In the IV stage we add to the root of the verb which we have obtained, the "regular" ending (if necessary) corresponding to the conditions explained in the preceding stages of the working of the scheme.

II Scheme "Adjective", the second scheme of Synthesis in order of operation also begins with determination of the group of words which have to be varied in this scheme. These words are adjectives, including participles which replace the feature "verb" by "adjective" after treatment under the Synthesising scheme "verb", together with nouns and numerals which decline like adjectives (*i.e.*, take the same case endings as adjectives. Not only nouns like "ucheny (scientist),

* After unification of the verb endings in relation to the particle -SYA.

† In Synthesis schemes in all cases the regularities in the changes in the written but not in the oral form of the word are established since the system of machine translation suggested is intended for translation of written texts.

‡ In particular for groups of verbs combined into the divisions:—

NV—"irregular" formation in present tense.

PV—"irregular" formation in past tense.

PN—"irregular" formation in imperative mood.

PP—"irregular" formation in past participle.

§ Explanation of the symbols used in denoting the groups has been given above.

peremennaya (variable), postoyannaya (constant), &c., but also pronouns "OH (he) (ONA, ONO, ONI), KTO, CHTO (who, what), TOT, ETOT (this, that); TE, ETI, SAM (those, these, itself)* or numerals "ODIN, DVA, TRI, CHETYRE," numeral pronouns "MNOGO (many), NESKOL'KO (several), OBA (both)" and some others.

From among the words treated by the scheme we exclude those having the feature "invariable." A case of invariability of a word is also that in which the required word form corresponds to the dictionary form of the word. By means of a series of special checks all these cases are excluded from further processing by the scheme.

The feature "short form" of adjective is the first obtained by the scheme. For this purpose we first obtain the root of the word, to which, depending on the features of number and gender and also on the type of root, is added one of the six endings (A, YA, O, E, Y, I) or—as a special case—no ending at all.

The root of the word is obtained by simple elimination of the last two letters (or four in the case of participles with the particle -SYA) from the dictionary form of the adjective. The only exception is in the group in which the suffix OV is added in forming the root of the short form of the adjective (KAKOY-KAKOV, TAKOY-TAKOV (such)).

For the short form of adjectives of masculine gender singular the possibility is taken into account of the fleeting vowel O (KRATKY-KRATOK (short) or E (IZVESTNY-IZVESTEN, VOZMOZHNY-VOZMOZHEN). This special feature of adjectival root formation is determined by referring them to the appropriate groups.

A second determination which precedes the actual variation of words by case, is connected with obtaining the feature "ordinal adjective" in those cases where it is subject to word variation. For this purpose the ending -Y or -OY is added to the original numeral, to achieve the possibility of alteration (of letter order) (e.g., VOSEM'-VOS'M + OY, SEM' SED'M + OY) or addition (e.g., SOROK-SOROK + OV + OY) in the root being allowed for.

At the beginning of this part of the scheme when the required case endings are found, words with "irregular" case endings (section L groups) are removed. These include pronomial adjectives and nouns (VES' (all), NASH (our), SAM (self), TOT, ETOT, TE, ETI, KTO (who), CHTO (which), ON (HE)†.

Determination of the case endings of a word is carried out separately for words with the feature "sing. no." and for words with the feature "plur. no."

In the first case it is established by a checking system whether irregular endings are required (for pronouns with the feature "irregular," these having been enumerated above) -O/E, A/YA, U/YU‡ and I the irregular ending -EM in the instrumental case (VES', TOT and the KTO-CHTO group). All remaining cases of cases change in the word are dealt with in the same way by the scheme "regular" word-variation: by rejecting the last two letters (in a special case—four letters§) from the dictionary form of the word and depending on the results of checking on the features of case and quality of the root one of the regular endings is added: AYA/YAYA, -UYU/YUYU, OY/EY for words of feminine gender and OYE/EYE for words of neuter gender; for words of masculine gender and in the remaining oblique cases of neuter words -OGO/EGO, -OMU/EMU, YM/IM, OM/EM.

Among words with the feature "plural number," those which are irregular are again separated, these being dealt with first, and also the "regular" forms of "irregular" words. In forming the root, in these cases as in the first, the special features of individual groups of words is taken into account, particularly "irregular" pronouns.

After joining the case ending to the root of an adjectival word with the feature "participle" the particle -SYA is restored if it was removed at the beginning of

* The words TOT, ETOT, TE, ETI, SAM are described in our MT dictionary as "adjectives" but in the Analysis schemes this description can be changed to "noun" if the adjective is substantivised in the sentence.

† Pronouns ONO, ONA, ONI [it, she, they] are regarded as word-forms of ON words derived from KTO and CHTO (NIKTO, NICHTO, NEKTO, NECHTO) and also KTO and CHTO with particles -TO, -LIBO, -NIBUD', KOYE- are included with them in one group [some, any-, whatever, &c.].

‡ The selection of one of two endings in this case is determined by a special check relative to the type of root.

§ The nominative and accusative for inanimate things of the masculine gender coincide with the dictionary form of the word and are excluded from the working of the scheme at the very start (see above).

the scheme. In a similar manner we remove at the beginning of the scheme (and restore at the end) the particles -TO, -LIBO, -NIBUD'.

III Scheme "Noun and Numeral" is the last in sequence in the scheme of Russian synthesis. At the beginning of the scheme we determine by certain checks (blocks 1-6) whether the word undergoes change in the scheme, *i.e.*, if it has the feature "noun"*, which includes "verbal noun", the part-of-speech feature of which was changed after treating in the synthesising scheme "verb" or "numeral" declining like a noun (*e.g.*, PYAT' (five), SHEST' (six) and other numeral ending m-': this does not apply to "irregular" numerals "SOROK (40), DENANOSTO (90), STO (100)" or to the numerals TYSYACHA (1,000), MILLION, MILLIARD†).

Words with the feature "invariable" are excluded from the working of this scheme and also words whose required word form coincides with the dictionary form. This includes cases where the grammatical characteristic of the word has the feature "nominative case" combined with the features "numeral" or "pronoun"; for other words the feature "singular no." is required as an additional characteristic. Besides this nouns in the singular do not change in accordance with the scheme if they have the features "accusative case" + "inanimate" combined‡.

Obtaining alternative grammatical features by the "noun" scheme commences with cases of irregular (exceptional) formation of pronouns§ and some numerals. These cases are solved individually. For other words the need for individual treatment arises only in determining the root of words complicated by transposition of letters. These words are subsequently submitted to a series of type checks by means of which the conditions for selecting the correct case ending become apparent.

For pronouns, three types of endings are distinguished and the second stage of the working of the scheme is divided into three parts corresponding to these.

In connexion with the special features of flexional formation of word forms, numerals which are declined like nouns receive a dictionary feature indicating the type of declension. Thus numerals ending in -' (PYAT', SHEST', DVADTSAT' (20), TRIDTSAT' (30), &c.), are given the feature "III declension" numerals TYSYACHA, MILLION, &c., respectively "II declension" and "I declension."

Nouns and numerals which form some case by a special "irregular" way are combined into groups and indications (by checking) are provided by the scheme. Thus for masculine nouns the regular form of the nominative plural is regarded as obtained by adding the termination Y/I to the root (STOL/STOLY (table); KOREN' -KORNI). Thus nouns forming the plural by adding the ending A (DOM/DOMA (house)) are regarded as exceptions and combined in a special group. Also in a special group are the masculine nouns used in the prepositional singular with the ending -U (instead of -E), *e.g.*, "V STOLE" but "V LESU" (in the wood) (Gr. IV VS-10K).

For nouns with the feature "plural no." the checking is considerably less owing to the fact that in the plural, differences in declension are found only in the nominative and genitive.

After the conditions for the selection of the case ending have been made clear, the required ending is added to the root in a uniform manner.

* Nouns which decline like adjectives do not fall under this scheme since they were treated in the preceding scheme.

† Numerals with double roots (PYAT' DESYAT [50], SHEST' DESYAT [60], DVESTI [200], &c.) are not varied according to the scheme but are written in figures.

‡ The feature "animate" is only given to words in the Russian MT dictionary where it is of grammatical significance, *i.e.*, only the group of masculine nouns, as the accusative case of these is of the same form as the genitive.

§ These include only the words Ya [I], MY [We], TY [Thou], VY [You], not having the form of the nominative SEBYA.

APPENDIX

Test translations. (Translator's comment: The translation is mainly correct and it seems rather pointless to reproduce it *in toto*. Instead some comments on the Russian rendering follow:)

General: The Russian text is not exactly what a Russian would produce, but it would be intelligible and convey the correct sense. It could be like English written by a foreigner, to some extent, to us.

Tests 2300-04: Has there been "cheating" in the form of touching up? It seems unlikely that "mass produced", "price" would be included in a mathematical dictionary? Perhaps not. And the word for "tackled"—quite correct, but "tackled" is not a very common English word in scientific texts. Grammatical error in Russian (ANALITICHESKOY RESHENIYE should read ANALITICHESKOGO RESHENIYA in a negative sentence: "analytic solution" in English). The last sentence would probably be rendered in "live" Russian by "staff without mathematical qualifications."

Tests 2400-07: "Imeyet mesto" means "takes place"; "occur" would be more usually rendered by words meaning "are found" or "are of seldom occurrence"; "lozhno polozheniya" means "false assumption"; "trial and error" would be better rendered by the phrase literally meaning "successive approximation" and this verges on wrong translation.

Tests 2311-25: The word for "has a knowledge of" would not be used naturally: it would be "is acquainted with" but this could be excused on the grounds of limiting the vocabulary. The translations in paragraphs headed 1 to 5 seem very good, presumably because concerned with technical mathematical terms. Doubtful if "economical" would be rendered literally: in Russian it normally means literally pertaining to money values and the word rendering "engineering" is unusual especially as the terms for the divisions of engineering are the normal ones. This and the absence of a rendering for "presentation" and even more surprisingly "space" indicates curious gaps in the vocabulary used. The rendering of "as wide as possible" is not good Russian. In the next sentence "familiar with" is the phrase which would naturally have been used in the first sentence of this test. It is also odd that "efficient" can be translated but not "efficiency." The Russian of the sentence beginning "The numerical tools" is awkward but would probably be intelligible: this sentence is one which would be expressed rather radically differently by a Russian and the same applies to the next sentence: the snag is "applied" (incidentally does the English really mean "applied to" or should it be "applied by" or "carried out by").

Summary: Most of the above comments relate to style and are perhaps perfectionist. There is (as far as I can detect) only one mistake in Russian grammar ("analytic solution") and one mistranslation ("false assumption"). The remaining comments mainly relate to what is in and what is omitted from the vocabulary. The general impression is one of considerable success; it is a notable achievement to be able to translate so well by a machine.